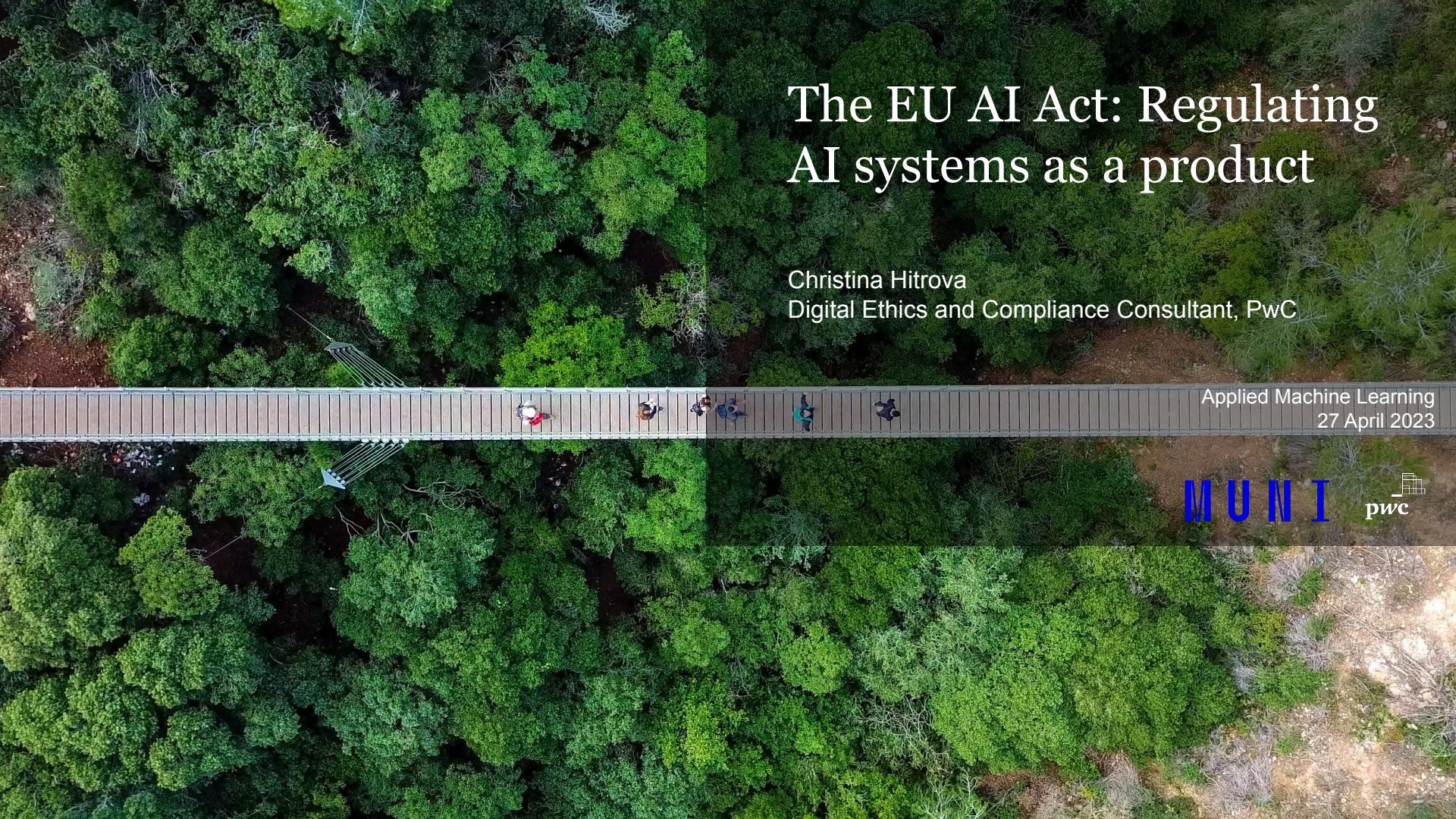


The EU AI Act: Regulating AI systems as a product

Christina Hitrova
Digital Ethics and Compliance Consultant, PwC

Applied Machine Learning
27 April 2023

MUNI

pwc

Christina Hitrova
Consultant in Digital Ethics and Compliance

christina.hitrova@pwc.com

Education

International and European Law

university of groningen

KU LEUVEN

University of Zurich[UZH]

Career

TRILATERAL RESEARCH

Data protection, privacy-by-design, civil drones, blockchain

The Alan Turing Institute

Ethics of data science in the UK public sector, AI ethics, digital welfare state

Responsible innovation, law and technology

pwc

Helping large organisations and their AI innovation comply with legal and ethical requirements

What kind of questions arise in the fields of law, ethics and technology?

Ethics - voluntary

What **ethical principles** can guide our actions to ensure a better outcome?

What **impact** does this technology and the way it is being used have on individuals and society?

What are the **reasons** for this impact? *(design, features, technical limitations, human use)*

How can we design or use technology in a way that

- Maximises positive effect
- Minimises negative effect

Law and regulation - mandatory

What **laws and regulations** apply to us and what do we have to comply with?

How do we translate **legal requirements into technical requirements?**

How does our technology, the way we design, sell and use it, measure against the legal requirements? Are there any **gaps** we need to fix?

How can we implement legal requirements in practice and **create an auditable record** to prove we comply?

Is AI good or bad? Both or neither?

Should we regulate AI technology? Why? When?

Should we regulate AI technology? Why? When?

FREE MARKET APPROACH / SOFT LAW & ETHICS

Rational market actors and market dynamics are enough to ensure that poorly created or harmful AI systems are driven out of the market. Therefore, it is sufficient to let the industry “self regulate”, potentially through non-binding guidelines.

Advantages

- **Does not hinder innovation** because allows complete freedom to all market actors
- Allows **a period of experimentation** to understand and learn about the technology
- It is **more flexible,** can be quickly changed

Disadvantages

- **Not binding** and cannot be enforced
- **There may be market failures,** e.g. market actors do not always have the information and expertise to act rationally due to information asymmetries between the creators of AI and other stakeholders

REGULATION AND HARD LAW

Binding legislation and regulation should be introduced to tackle issues that the free market of rationally acting supply and demand does not resolve (market failures).

Advantages

- **Binding** and can be legally enforced
- **Limits innovation** by prohibiting or shaping it through requirements
- **Tackles market failures**, e.g. by shifting the risks of AI innovation onto AI creators or users or by increasing information on the market

Disadvantages

- Requires a **long legislative process** which can make it outdated
- May **hinder innovation** through high compliance costs
- **May be premature** if regulate without sufficient understanding of the technology

Are there any market failures?

A tech bro creates an AI to **automate review of health insurance claims** and identify which claims should be successful, given historical data. It is a **continuous learning** application.

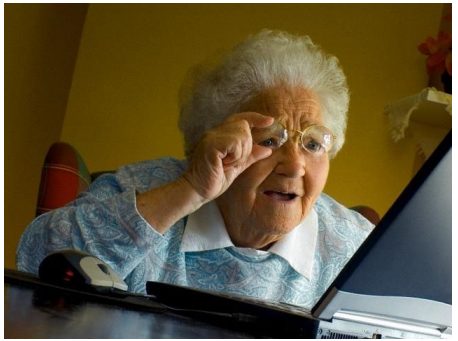
He wants to keep his work a **trade secret** because it is so valuable.

An **insurance company** buys the AI tool to use it with its customers wit**hout understanding** how it works.

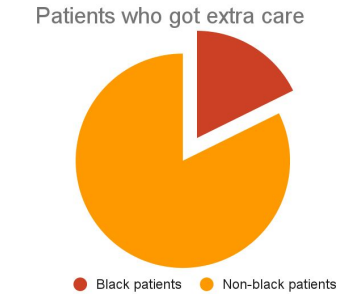
An employee runs the tool on the claim by an elderly lady and gets a negative result. **She refuses to pay the claim** and doesn't give further explanation except that their AI system said so.

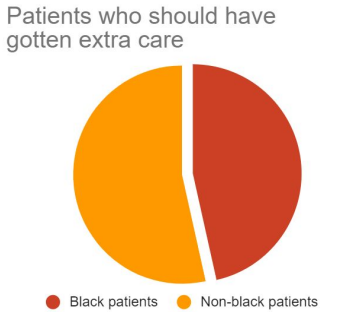
The **elderly lady gets a rejection** from the insurance company.

She gets **no reasoning** from the company and **cannot access or understand the AI system**.



Millions of black people affected by racial bias in health-care algorithms





An algorithm used to allocate health care to patients based on need was consistently biased against black people

Why did this happen?

Expected healthcare costs ≠ Healthcare needs

Historically, black patients had less access to healthcare insurance and services in the US. **The algorithm perpetuated this pattern and prevented access to healthcare for millions who needed it.**

Ledford, Nature 574, 608-609 (2019).

Some challenges in the context of AI systems

Information asymmetry

Information and know-how about AI systems - how to create, use, and understand them - is not equal among all market actors and may sometimes even be protected intellectual property.

People don't always understand what they use, how to choose the best product or maintain it.

Many hands problem

Many different teams or companies can be involved in the creation of 1 AI tool. This becomes more complex if using open source code or foundational AI models.

If something goes wrong, who is responsible?

Foreseeability and autonomy

AI systems can operate with autonomy, including continuous self-learning. Humans cannot foresee how these systems will operate in the future.

If something goes wrong, who should be responsible?

Explanation and remedies

Some AI systems are so complex that not even their creators understand how they operate and why a specific result was achieved.

If we don't understand AI systems, how to determine whether something went wrong and who is responsible? How can affected individuals ask for a remedy?

Authority and scope of impact

Once created, AI systems present an "aura of objectivity" - it is a programme that everyday people cannot argue with. If it malfunctions, this can have a negative impact on a large scale.

How should we ensure AI is as safe as possible before we deploy it?

Context, relevance, bias

The quality of an AI system, like any statistical method, requires understanding the context, the available data, and the intended use case to make the best design choices possible.

What happens if we don't have the right amount and quality data available? What about data that is biased? What could the effects be?

Commonalities with other complex products like pharmaceuticals and financial investment products

Parallels with:

- Pharmaceuticals
- Financial investment products
- Aviation
- Civil engineering, architecture

Experts have knowledge that those affected do not. There is need for trust.

Unknown long-term impact of the product. Nothing is guaranteed.

Significant and large scale potential impact on people

Regulated professions of doctors, pharmacists

Regulated process of creating products, e.g. testing in pharmaceuticals

Requirements to know your customer (in finance) and **to provide information** to consumers, sometimes in language tailored to their experience and knowledge.

An overview of the European Union's draft AI Act

pwc

In April 2021 the European Commission unveiled its proposal for a Regulation on a European Approach to Artificial Intelligence.

The text of the Regulation is being negotiated by EU legislators, but the direction of change is clear.

Legislators want more safeguards and obligations for producers and users of AI systems. Consumers want higher quality and safer innovation.

AI Act goals - regulating AI as a product to ensure it is safe before and while on the market

Product requirements & standards

The AI Act prescribes a number of **concrete rules and requirements** that high-risk AI systems must achieve *before* being sold.

Standards will be developed to clarify how to comply with these requirements.

Conformity assessment and CE marking

Before placing it on the market, providers of AI systems must undergo a **conformity assessment** to confirm the AI systems meet the necessary requirements.

If they are successful, they can mark the AI system with the EU product CE marking.

Post market monitoring

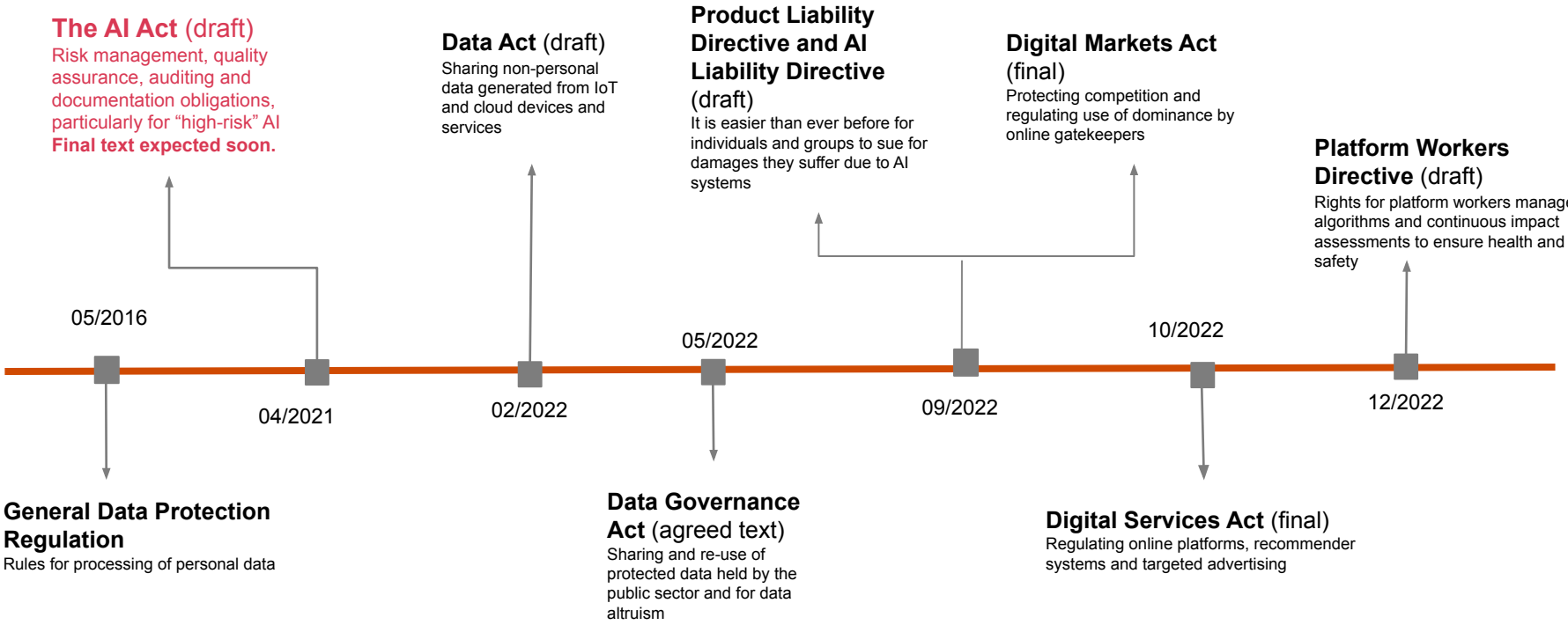
After being placed on the market, providers **have to keep monitoring the performance** of the AI system and have to notify and take action to remedy any unexpected behaviour.

Responsibility and liability of producers

In a different proposed law, **the Product Liability Directive,** providers of AI systems as products are held responsible for material harms their products caused when used as intended.

This is a **strict liability** meaning they are responsible regardless of whether they did something wrong (negligent, reckless).

The AI Act is one of many legislative initiatives to govern the EU AI and data space

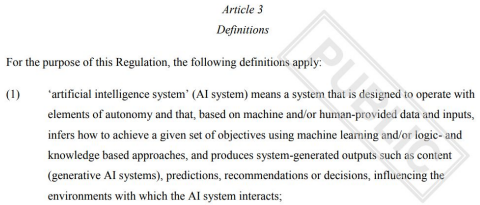


Who will be affected by the AI Act?

Scope

Extra-territorial scope: anyone based in the EU or wanting to offer an AI system to residents within the EU

Broad definition of AI that covers more than just ML



Article 3

Definitions

For the purpose of this Regulation, the following definitions apply:

(1) 'artificial intelligence system' (AI system) means a system that is designed to operate with elements of autonomy and that, based on machine and/or human-provided data and inputs, infers how to achieve a given set of objectives using machine learning and/or logic- and knowledge based approaches, and produces system-generated outputs such as content (generative AI systems), predictions, recommendations or decisions, influencing the environments with which the AI system interacts;

Source - Council version of AI Act

Providers vs users

Most requirements to ensure AI is safe fall on providers.

Providers: Those who

- Create AI systems for themselves or others
- Repurpose existing AI systems
- Import and distribute AI systems from outside the EU

Users: Those who just use AI systems that already exist.

Risk-based approach

The AI Act obligations can be difficult to comply with so the law focuses its heaviest requirements on the riskiest AI systems.

Obligations are focused on specific use cases of AI systems that can have a potential impact on the health, safety and fundamental rights of individuals.

Do you have ideas what this could be?

The AI Act's risk-based approach: How the AI Act will affect you depends on the AI use cases you produce and use

The scope of the AI Act is very broad. AI systems that are produced or sold in the EU or to EU residents are covered. The definition of an AI system is equally wide and covers not only advanced analytics techniques but also more basic software. Existing and new AI systems are affected and there are obligations on both producers and users of such technology.

Unacceptable risk

- Social scoring
- Trustworthiness of individuals
- Real-time remote biometric identification
- Harmful subliminal techniques

Such solutions, as defined in the final AI Act, will be **prohibited** for placing on the market in the EU.

High risk

Industry angle:
Stand-alone AI product, regulated by product safety regulations

AI system as a safety component of regulated product

Use case angle:
A list of predetermined AI use cases with an impact on the health, safety and fundamental rights of individuals

The way such solutions are produced, assessed for compliance, placed on the market, used and maintained is **regulated** by the AI Act.

Limited risk

- Chatbots
- Biometric categorisation
- Media generation, "deepfakes"
- Emotion recognition

These systems are permitted, subject to specific transparency obligations to inform affected persons.

Low and no risk

All other systems are permitted without any additional restrictions. They may, however, comply with the AI Act *voluntarily*.

What is an AI system?

- designed to operate with elements of **autonomy**
- based on machine and/or human provided data and **inputs**
- **infers** how to achieve a given set of objectives
- using machine learning and/or logic- and knowledge based approaches
- produces system-generated **outputs** such as content (generative AI systems), predictions, recommendations or decisions
- **influencing** the environments with which the AI system interact

High-risk AI systems are heavily regulated in the AI Act. There are two types of high-risk AI systems

1 Industry angle

Stand-alone products or a safety component of a product that is regulated and has to undergo a third-party conformity assessment, including:

Industry
Machinery, equipment for potentially explosive atmospheres, burning gaseous fuels, pressure equipment

Healthcare
Medical devices, in vitro diagnostic devices, personal protective equipment

Transportation
Motor vehicles, civil aviation, rail system, marine equipment, agricultural and forestry vehicles, 2+ wheeled vehicles

2 Use-case angle

Specified use cases, including:

Recruitment and employee management

Biometrics
identification or categorisation of persons

Critical infrastructure
Operation, safety, environmental emissions, pollution, control

Finance and insurance
Insurance premiums setting
Creditworthiness assessments

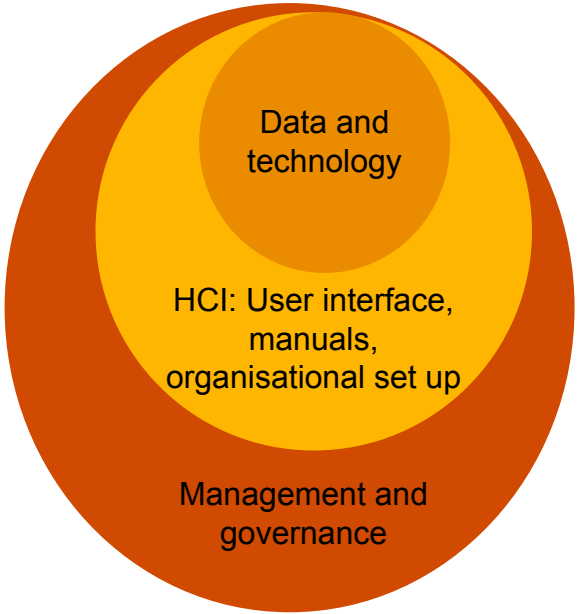
Education and training

Administration of justice and democratic processes

Public sector
Social welfare, law enforcement, emergency services; border control and migration policy

* The text of the AI Act is still being negotiated and may change in the future.

The AI Act requirements ensure high-risk AI systems are safe before and while on the market



Data quality requirements and bias checks

Accuracy, robustness & cybersecurity

Record-keeping and logs of operation

User transparency, capability and maintenance

Human oversight and control

Quality management system

Risk management system

Conformity assessment

Data management and governance

Technical documentation

Post market monitoring

* The text of the AI Act is still being negotiated and may change in the future.

Compliance by design

"By design" refers to the process of designing a data-driven or software solution in a way that it meets certain criteria and requirements or achieves specific goals. This requires:

- Clear requirements to be achieved
- Continuous assessment of proposed designs against requirements
- Iterative tackling of gaps to achieve requirements

Privacy-by-design
Data protection-by-design
Security-by-design

Ethics-by-design

Compliance-by-design

Requirements on data and technology

Requirements on data quality and bias (Art. 10)

Training, validation and testing datasets should be

- Relevant
- Representative
- As far as possible free from errors and complete

Data should be examined for biases that can affect health and safety or lead to discrimination.

You must consider the **precise context** where the systems will be used (or foreseeably will be used).

How to get data about statistical distribution in reality to assess representativeness? How to get personal data to assess biases?

Accuracy, robustness, cybersecurity (Art. 15)

Appropriate levels of accuracy, robustness and cybersecurity are consistently needed.

The accuracy levels and metrics should be clearly declared.

The risk of feedback loops in continuous learning tools should be addressed,as should be the risk from cyber and adversarial attacks.

What does “appropriate” levels of accuracy for different intended uses mean? What kind of accuracy and along what metrics should we assess? How to control for feedback loops when users use AI separately?

Record-keeping and logging (Art. 12)

The system should allow for the automatic recording of events over its lifecycle - how it performs overtime, its operation, etc.

This is also a key feature to allow for *post market monitoring.*

What should be recorded about the AI system’s operation? How should providers best set up the collection of data from their users about the AI system’s performance?

Requirements on human-AI interaction, use, and maintenance

User transparency and information (Art. 13)

System operation should be sufficiently transparent to ensure users can understand and use the system's output appropriately.

Systems should be accompanied by instructions that includes:

- characteristics, capabilities and limitations of performance of high-risk AI system
- appropriate human oversight
- running and maintenance requirements of the system

What is sufficiently transparent? What degree of explainability might be needed and can it be achieved? How should instructions be formulated to be clear? For what kind of user?

Human oversight and control (Art. 14)

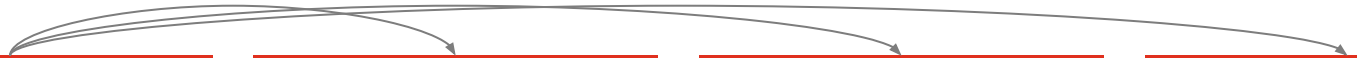
The system should be designed and have appropriate interface to be effectively overseen by humans during use.

Individuals responsible for human oversight should

- have sufficient understanding of the system
- understand automation bias
- be able to interpret the system's operation
- be able to stop the system at any time.

What should an interface look like and what features should it have to enable human oversight? How should foreseeable misuse be defined and determined? What guidance for deploying AI systems and surrounding procedures should be given to users?

Requirements on setting up processes and procedures



Quality management system (Art. 17)

Providers should set up written policies, procedures and instructions to ensure the AI systems they create comply with the AI Act requirements. This should be proportionate to their size and cover *inter alia* techniques, procedures and actions to design, control the quality of AI, validate and test AI at different stages of design and roles and responsibilities.

What factors to consider to determine proportionality of QM to size of provider? Can we have a taxonomy of best practices for actions to take during design or quality assessment? What quality features to especially focus on?

Risk management system (Art. 9)

Providers should have a system to iteratively identify, analyse, eliminate and mitigate known and foreseeable risks to health, safety and fundamental rights. Risk mitigation steps must be taken continuously to reach “acceptable” risk levels. AI systems should be tested to assess their risk and ensure they perform as intended.

Can we have a taxonomy of known and foreseeable risks per high-risk use case and their underlying causes? Can we quickly recommend risk mitigation? What is “acceptable” remaining risk?

Data management and governance (Art. 10)

Training, validation and testing data should have appropriate data management that cover:

- Relevant design choices
- Data collection
- Data preparation
- Formulation of relevant assumptions wrt the data and its meaning
- Assessing availability, suitability and quantity of data
- Data biases
- Data shortcomings and how to address them

What are “appropriate” data governance and management practices? What are the features of data management necessary to prevent AI risks? What does “relevant design choices” mean? Can any of this be automated?

Post market monitoring (Art. 61)

Providers should have a post-market monitoring system that is *proportionate* to the risk of the AI. The system should actively and systematically collect, document and analyse data from users or other means about the performance of the AI throughout its lifetime that allows the provider to evaluate its compliance with the AI Act.

What comprises a post-market monitoring plan? What is proportionate to different high-risk AI use cases and what factors to consider? What information should be collected to assess compliance continuously? How to best collect it?

The approval process before entering the EU market

Technical documentation (Art. 11, Annex IV)

A set of standardised technical documentation should be drawn up and maintained. It is reviewed during the conformity assessment and demonstrates the AI's legal complinace.

The documentation includes a description of the AI system and how to use it, how it operates and how it was designed, its design specifications, its validation and testing…

Conformity assessment (Annex VII)

A conformity asssesment process checks whether a product complies with the relevant requirements.

High-risk AI systems have undergo through a conformity assessment.

- Mostly *internal* CA
- A third party CA s required for specific cases, e.g. medical devices, or AI as part of regulated products like cars.

Public database (Art. 61, Annex VIII)

A public register, maintained by the European Commission, will have a list of all *stand alone* high-risk AI systems on the EU market.

It includes who is responsible for the AI system, what is its intended use, status and instructions for its use.

Two paths to compliance

Definition

Do it yourself	Rely on harmonized standards
Use your best judgment about what the text of the AI Act means and set up your own processes to comply: • Quality management system • Risk management system • Data quality and data management • Accuracy, robustness, cybersecurity • …..	Rely on *harmonized standards* to be developed by CEN/CENELEC that clarify the AI Act requirements. • Automatic presumption of compliance with the law • Developed by technical experts, not law-makers

How do you assess the AI Act?

FREE MARKET APPROACH / SOFT LAW & ETHICS

Rational market actors and market dynamics are enough to ensure that poorly created or harmful AI systems are driven out of the market. Therefore, it is sufficient to let the industry “self regulate”, potentially through non-binding guidelines.

Advantages

- **Does not hinder innovation** because allows complete freedom to all market actors
- Allows **a period of experimentation** to understand and learn about the technology
- It is **more flexible,** can be quickly changed

Disadvantages

- **Not binding** and cannot be enforced
- **There may be market failures,** e.g. market actors do not always have the information and expertise to act rationally due to information asymmetries between the creators of AI and other stakeholders

REGULATION AND HARD LAW

Binding legislation and regulation should be introduced to tackle issues that the free market of rationally acting supply and demand does not resolve (market failures).

Advantages

- **Binding** and can be legally enforced
- **Limits innovation** by prohibiting or shaping it through requirements
- **Tackles market failures**, e.g. by shifting the risks of AI innovation onto AI creators or users or by increasing information on the market

Disadvantages

- Requires a **long legislative process** which can make it outdated
- May **hinder innovation** through high compliance costs
- **May be premature** if regulate without sufficient understanding of the technology

How would the AI Act change this situation?

A tech bro creates an AI to **automate review of health insurance claims** and identify which claims should be successful, given historical data. It is a **continuous learning** application.

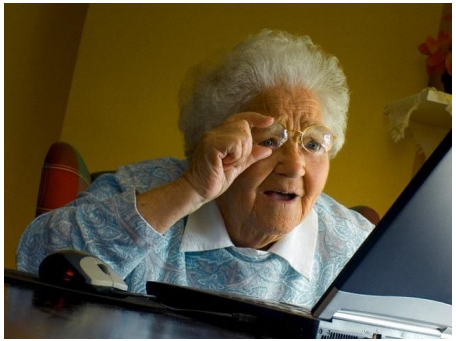
He wants to keep his work a **trade secret** because it is so valuable.

An **insurance company** buys the AI tool to use it with its customers wit**hout understanding** how it works.

An employee runs the tool on the claim by an elderly lady and gets a negative result. **She refuses to pay the claim** and doesn't give further explanation except that their AI system said so.

The **elderly lady gets a rejection** from the insurance company.

She gets **no reasoning** from the company and **cannot access or understand the AI system**.



Is the AI Act premature? How is ChatGPT changing it?

The AI Act - what does it mean for foundational models or general purpose AI like chatGPT?

In or out of AI Act scope?

Council: GPAI that has an intended use or foreseeable intended use that is high risk to be high-risk AI

EP: Generating complex text without human oversight to be “high-risk”?

The risk-based approach: Classifying the risk of GPAI isn’t easy. But Big Tech promotes maintaining risk-based AIA approach and keeping GPAI out of it.

Politico discussion

Are AI Act requirements enough for ChatGPT risks?

Halucinating and lack of sources

Malicious use, e.g. social engineering, hacking

Privacy

Discrimination and stereotyping

Human overreliance and anthropomorphising

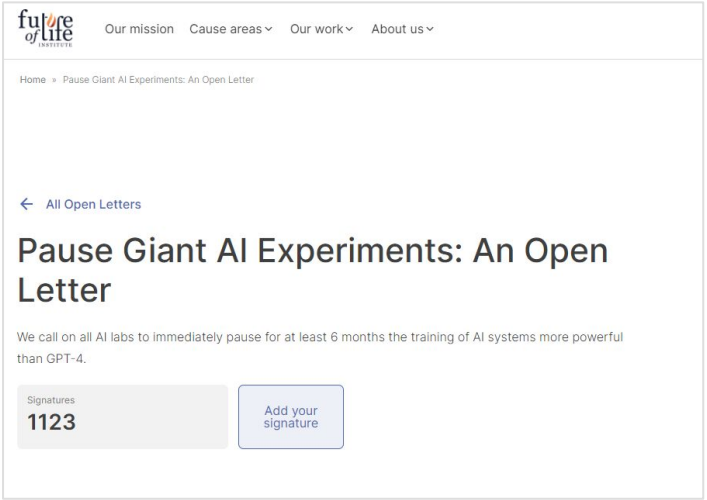
Increasing economic inequality and disruption

Environmental costs

Built on knowledge by all, owned by a handful few

Synthesising vs generating information - Shahab Anbarjafari’s post

Future of Life Institute open letter calls for a **moratorium on new LLMs**



Link

Christina Hitrova

christina.hitrova@pwc.com